Wandering in the Labyrinth of Thinking
– a cognitive architecture combining reinforcement learning and deep learning

甄景贤 (King-Yin Yan) and Juan Carlos Kuri Pinto

General.Intelligence@Gmail.com

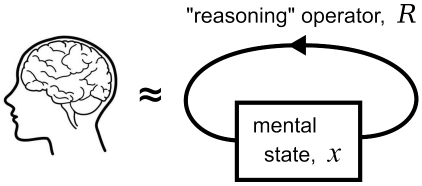
Abstract. The problem of general intelligence can be described and solved in the logic-based AI paradigm, but the main obstacle is that learning is too slow. The logic-based knowledge representation with discrete propositions is abandoned in favor of a neural-based "amorphous" representation induced from the top down, using a deep neural network (DNN). The DNN acts iteratively on a state space (the "mental state"), forming a dynamical system. This system is in turn controlled by reinforcement learning, "navigating" the space of mental states as in a maze.

1 Main idea

The **metaphor** here is that of reinforcement learning controlling an autonomous agent to navigate the maze of "thoughts space":

(1)

The main idea is to regard "thinking" as a **dynamical system** operating on **mental states**:



(2)

For example, a mental state could be the following set of propositions:

- I am in my room, writing a paper for AGI-16.

– I am in the midst of writing the sentence, “I am in my room, ...”
– I am about to write a gerund phrase “writing a paper...”

Thinking is the process of **transitioning** from one mental state to another. Even as I am speaking now, I use my mental state to keep track of where I am at within the sentence’s syntax, so that I can structure sentences grammatically.

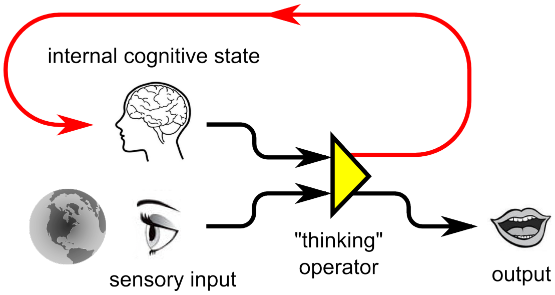
By representing a cognitive state as a vector $\vec{x} \in X$ where X is the cognitive state-space, the reasoning operator R as an **iterative map** $X \to X$, we would have at disposal all the tools available in vector space such as:

- numerical optimization (eg gradient descent)
- differential equations governing time evolution
- dynamical systems theory, control theory, dynamic programming, reinforcement learning
- neural networks and deep learning ... etc.

1.1 Related work

Google’s **PageRank** is one of the earlier successful applications of vector-space and matrix techniques. The **Word2Vec** [4] algorithm that maps natural-language words to vectors is also spectacularly successful and influential; it demonstrated the potential advantages of vector representations. As for reinforcement learning, Q-learning (a form of RL) has been combined with deep learning to successfully play Atari games [2]; Their architecture is exactly the same as ours, except that we are trying to refine the internal structure of the learner.

This is the cartoon version of our architecture:



(3)

2 Logic-based AI (LBAI)

Main points:

- We would not directly implment logic-based AI, but it serves as a *backdrop* for understanding what are the problems of general AI.
- In this paper we would jump back and forth between the logic-based view and the dynamical state-space view. Knowledge of LBAI is essential to understanding ideas in this paper.

It is feasible to use mathematical logic to emulate human thinking, an approach pioneered by John McCarthy (1927-2011). We have 3 basic operations: deduction, abduction, induction; For details one can refer to《Computational logic and human thinking》by Robert Kowalski, 2011. We would not waste time to debate whether LBAI is an adequate model of human thinking; This paper assumes it as the point of departure. It is worth mentioning though, that Kowalski is one of the researchers who laid the theoretical foundations of logic programming, especially Prolog.

In classical logic-based AI, "thinking" is achieved by steps like this:

$$\text{premise} \vdash \text{conclusion} \tag{4}$$

$$\boxed{\text{it was raining this morning}} \vdash \boxed{\text{grass is wet}} \tag{5}$$

That is to say: from some **propositions** we deduce other propositions.

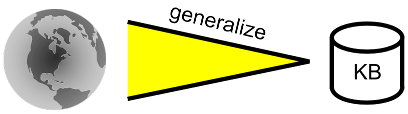
Deduction requires some special propositions known as **rules**, these are propositions containing **variables** such as "x":

$$\boxed{\text{it is raining at location } x} \wedge \boxed{x \text{ is uncovered}} \vdash \boxed{\text{location } x \text{ is wet}} \tag{6}$$

Rules are like the "fuel" for an inference engine; The engine cannot run without fuel.

Note: The x inside a proposition is like a "hole" in it. We could use **substitution** to place some concrete **objects** into such holes, to make the proposition *complete*. This is a form of **sub-propositional** structure, and one way to express it is via **predicate logic**. We don't need to concern with details right now.

LBAI can be viewed as the compression of a world model into a knowledge-base (KB) of logic formulas (that consists of **facts** as well as **rules**):

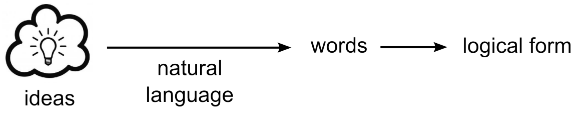


(7)

The world model is *generated* combinatorially from the set of logic formulas, vaguely reminiscent of a "basis" in vector space. The generative process in logic is much more complicated, contributing to its high *compressive* ability on the one hand, and the *complexity* of learning such formulas on the other hand.

2.1 Bottom-up vs top-down representations

In LBAI the knowledge representation structure is built (*fixed*) from the bottom up (For example, predicate symbols and constant symbols build up propositions, and sets of propositions form theories):



(8)

but is it valid (or profitable) to assume that our mental representations are *isomorphic* to such logical structures? Or drastically different?

The most serious disadvantage of bottom-up representations lies in the difference between **syntactic distance** and **semantic distance**. Suppose propositions are built up from an "alphabet" of atomic concepts, through the use of a multiplication operation such as tensor product. We embed atomic concepts into a vector space, in the manner of the Word2Vec algorithm. Then, using the tensor product, propositions (ie sentences) will be mapped to positions in the tensor-product vector space. Thus we can measure the **distance** between any two propositions. However, this is a **syntactic** distance. For example, *"Don't judge a book by its cover"* and *"Clothes do not make the man"* are superficially very different (syntactically distant) but are semantically close. In a good learning system we need to **generalize** according to semantic distance. The embedding of bottom-up representations usually gives us a discrete space with fractal structure, and the metric defined on such a space is always syntactic.

Humans are good at designing symbolic structures, but we don't know how to design *neural* representations which are more or less opaque to us. Perhaps we could use a neural network acting recurrently on the state vector to **induce** an internal representation of mental space. "*Induced by what*," you ask? By the very structure of the neural network itself. In other words, forcing a neural network to *approximate* the ideal operator R^*.

From an abstract point of view, we require:

- R be an endomorphism: $X \to X$
- R has a learning algorithm: $R \overset{A}{\longmapsto} R^*$

R would contain all the knowledge of the KB, so we expect it to be "large" (eg. having a huge number of parameters). We also desire R to possess a **hierarchical** structure because hierarchies are computationally very efficient. A multi-layer perceptron (MLP) seems to be a good candidate, as it is just a bunch of numbers (weight matrices W) interleaved by non-linear activation functions:

$$R(\boldsymbol{x}) = \sigma(W_1 \sigma(W_2 ... \sigma(W_L \boldsymbol{x}))) \tag{9}$$

where L is the number of layers. MLPs would be our starting point to explore more design options.

In 1991 Siegelmann and Sontag [3] proved that recurrent neural networks (RNNs) can emulate any Turing machine. In 1993 James Lo [1] proved that RNNs can universally approximate any non-linear dynamical system.

The idea of R as an operator acting on the state is inspired by the "consequence operator" in logic, usually denoted as Cn:

$$\mathrm{Cn}(\Gamma) = \{ \text{ set of propositions that entails from } \Gamma \ \} \tag{10}$$

but the function of R can be broader than logical entailment. We could use R to perform the following functions which are central to LBAI:

- **deduction** – forward- and backward-chaining
- **abduction** – finding explanations
- **inductive learning**

Below, we try to formalize the structure of logic from 2 perspectives:

- Static structure (formulas built from atomic concepts, logic operators, etc)
- Dynamic structure (mechanisms of proof, inference, etc)

2.2 Static structure of logic

- **truth values** (eg. P(rain tomorrow) = 0.7)
- **propositional structure** (eg. conjunction: $A \wedge B$)
- **sub-propositional structure** (eg. predication: loves(john, mary))
- **subsumption structure** (eg. dog $\subseteq$ animal)

These structures can be "transplanted" to the vector space X via:

- **truth values:** an extra dimension conveying the "strength" of states

- **propositional structure:** eg. conjunction as vector addition,

$$A \wedge B \quad \Leftrightarrow \quad \boldsymbol{x}_A + \boldsymbol{x}_B + ... \tag{11}$$

  but we may have to avoid linear dependencies ("clashing") such as:

$$\boldsymbol{x}_3 = a_1 \boldsymbol{x}_1 + a_2 \boldsymbol{x}_2 \tag{12}$$

  This would force the vector space dimension to become very high.
- **sub-propositional structure:** eg. tensor products as composition of concept atoms:

$$\text{loves(john, pete)} \quad \Leftrightarrow \quad \overrightarrow{john} \otimes \overrightarrow{love} \otimes \overrightarrow{pete} \tag{13}$$

- **subsumption structure:** eg. define the **positive cone** C such that

$$\text{animal} \supseteq \text{dog} \quad \Leftrightarrow \quad \overrightarrow{animal} - \overrightarrow{dog} \in C \tag{14}$$

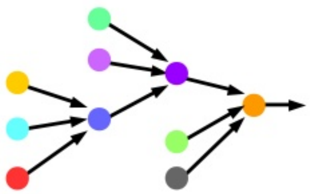
But the more logical structure we add to X, the more it will resemble logic, and this whole exercise becomes pointless. Remember our original goal is to try something different from logic, by *relaxing* what defines a logical structure. So we would selectively add features to X.

2.3 Dynamic structure of logic

@Andrew: This is where I want to formalize a "logical system". Particularly, the state X has internal structure that I have ignored so far: X should be a **set** of propositions. During deduction, we need to **select** a few propositions from X and try to **match** them with existing logic rules (this is the job of the famous **unifcation** algorithm in logical AI systems). The selection is part of the control variable u (see below). We need to decompose the vector X into some analogue of "propositions", but I don't know how to do it yet. Perhaps elucidating the algebraic form of the logic system will help us design the "vectorization" scheme.

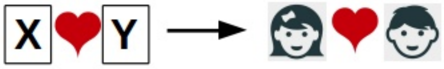
The 2 "pillar" algorithms for deduction in LBAI are:

- **Resolution**: deducing new propositions (conclusions) from existing ones (premises)



(15)

– **Unificaiton**: matching a proposition with variables (“holes”) with grounded (“without holes”) propositions



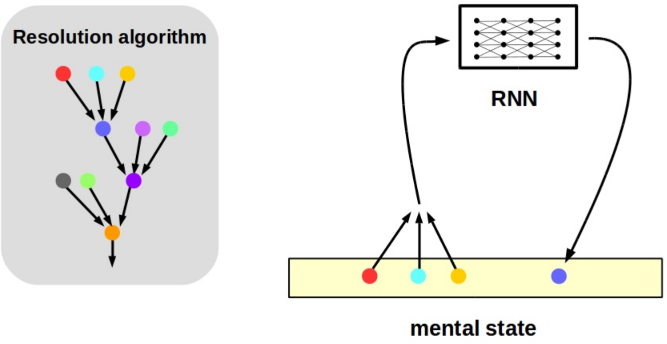
(16)

The algebraization of first-order predicate logic (a logic whose propositions can have internal variables) is a difficult subject, potentially involving Tarski’s cylindrical algebra which the author is unfamiliar with.

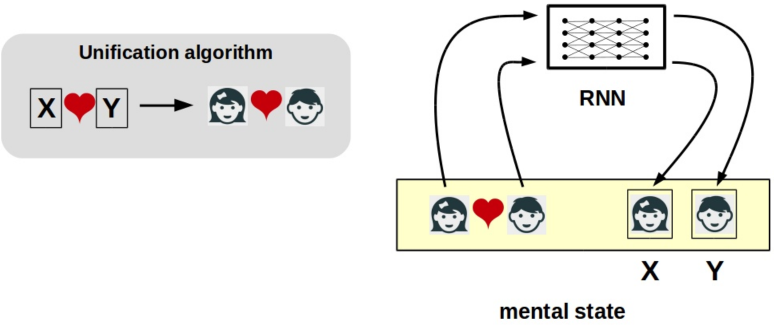
Here we introduce a crucial idea: using an **external memory** to manage the problem of **variable binding**. Recall that a **Turing machine** is just a finite state machine equipped with a **memory tape**; It could be said that the memory tape is what enables the machine to have Turing-complete computing power. Similarly, allowing a **propositional logic** to use an external memory storage for intermediate results, enables it to have the same expressive power as **predicate logic**.

Below are 2 cartoons illustrating how **resolution** and **unification** are performed with the aid of **external memory**:

@Andrew: I want to formalize these operations. It may be more important than formalizing the **static** properties of logic.



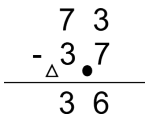
(17)



(18)

Example 1: primary-school arithmetic

A recurrent neural network is a *much more powerful* learning machine than a feed-forward network, even if they look the same superficially.



As an example, consider the way we perform 2-digit subtraction in primary school. This is done in two steps, and we put a dot on paper to mark “carry-over”.

The use of the paper is analogous to the “tape” in a Turing machine – the ability to use short-term memory allows us to perform much more complex mental tasks.

We did a simple experiment to train a neural network to perform primary-school subtraction. The operator is learned easily if we train the two steps *separately*. The challenge is to find an algorithm that can learn **multi-step** operations by itself.

Example 2: variable binding in predicate logic

The following formula in predicate logic defines the “grandfather” relation:

$$\text{grandfather}(X,Z) \longleftarrow \text{father}(X,Y) \wedge \text{father}(Y,Z) \tag{19}$$

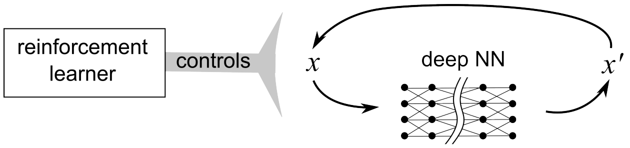
We did a simple experiment to train a neural network to perform primary-school subtraction. The operator is learned easily if we train the two steps *separately*. The challenge is to find an algorithm that can learn **multi-step** operations by itself.

3 Control theory / reinforcement learning

@Andrew: as you will see, the control theory part is essentially separated from the "logic" aspects.

Main points:

- Intelligence is decomposed into **thinking** and **learning**.
- **Thinking** is governed by control theory (finding the best trajectory in "thoughts space") under the contraints of correct reasoning, ie, knowledge.
- The iterative "thinking operator" is implemented as a deep-**learning** neural network (DNN). This DNN *contrains* the dynamics of thinking, and it represents the totality of *knowledge* in the system.



(20)

3.1 What is control theory?

A **dynamical system** can be defined by:

$$\text{discrete time:} \qquad \boldsymbol{x}_{t+1} = \boldsymbol{F}(\boldsymbol{x}_t) \tag{21}$$

$$\text{continuous time:} \qquad \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) \tag{22}$$

($\boldsymbol{F}$ is implemented as the deep learning network in our approach.)

A **control system** can be defined as (sometimes I mix with continuous-time notation for the sake of simplicity):

$$\dot{\boldsymbol{x}}(t) = f(\boldsymbol{x}(t), \boldsymbol{u}(t), t) \tag{23}$$

where $\boldsymbol{u}(t)$ is the **control vector**. The goal of control theory is to find the optimal $\boldsymbol{u}^*(t)$ function, such that the system moves from the initial state $\boldsymbol{x}_0$ to the terminal state $\boldsymbol{x}_\perp$.

3.2 What is reinforcement learning?

Reinforcement learning is synonymous with **dynamic programming**, which is also the main content of modern **control theory** with the state-space description.

The goal of **reinforcement learning** is to learn the **policy function**:

$$\text{policy}: \quad \text{state} \xmapsto{\text{action}} \text{state'} \tag{24}$$

when we are given the **state space**, **action space**, and **reward function**:

$$\text{reward}: \boxed{\text{state}} \times \boxed{\text{action}} \to \mathbb{R} \tag{25}$$

The action a is the same notion as the control variable u in control theory.

The **Bellman equation** governs reinforcement learning just as in control theory:

$$\begin{aligned} \boxed{\text{optimal path}} &= \text{choose max reward on current path segment} \\ &\qquad + \boxed{\text{the rest of optimal path}} \end{aligned} \tag{26}$$

In math notation:

$$U^*_t = \max_u \{ \boxed{\text{reward(u, t)}} + U^*_{t-1} \} \tag{27}$$

where U is the "long-term value" or **utility** of a path.

Conceptually, U is the **integration** of instantaneous rewards over time:

$$\boxed{\text{utility, or value U}} = \int \boxed{\text{reward R}} \; dt \tag{28}$$

3.3 Connection with Hamiltonian mechanics

An interesting insight from control theory is that our system is a Hamiltonian dynamical system in a broad sense.

Hamilton's **principle of least action** says that the trajectories of dynamical systems occuring in nature always choose to have their action S taking **stationary values** when compared to neighboring paths. The action is the time integral of the Lagrangian L:

$$\boxed{\text{Action S}} = \int \boxed{\text{Lagrangian L}} \; dt \tag{29}$$

From this we see that the Lagrangian corresponds to the instantaneous "rewards" of our system. It is perhaps not a coincidence that the Lagrangian has units of

energy, in accordance with the folk psychology notion of "positive energy" when we talk about desirable things.

The **Hamiltonian** H arises when we consider a typical control theory problem; The system is defined via:

$$\begin{aligned} \text{state equation:} \quad & \dot{\boldsymbol{x}}(t) = \boldsymbol{f}[\boldsymbol{x}(t), \boldsymbol{u}(t), t] & (30) \\ \text{boundary condition:} \quad & \boldsymbol{x}(t_0) = \boldsymbol{x}_0 \, , \, \boldsymbol{x}(t_\perp) = \boldsymbol{x}_\perp & (31) \\ \text{objective function:} \quad & J = \textstyle\int_{t_0}^{t_\perp} L[\boldsymbol{x}(t), \boldsymbol{u}(t), t] dt & (32) \end{aligned}$$

The goal is to find the optimal control $\boldsymbol{u}^*(t)$.

Now apply the technique of **Lagrange multipliers** for finding the maximum of a function, this leads to the new objective function:

$$U = \int_{t_0}^{t_\perp} \{L + \boldsymbol{\lambda}^T(t) \, [f(\boldsymbol{x}, \boldsymbol{u}, t) - \dot{\boldsymbol{x}}]\} dt \tag{33}$$

So we can introduce a new scalar function H, ie the Hamiltonian:

$$H(\boldsymbol{x}, \boldsymbol{u}, t) = L(\boldsymbol{x}, \boldsymbol{u}, t) + \boldsymbol{\lambda}^T(t) f(\boldsymbol{x}, \boldsymbol{u}, t) \tag{34}$$

Physically, the unit of $\boldsymbol{f}$ is velocity, while the unit of L is energy, therefore $\boldsymbol{\lambda}$ should have the unit of **momentum**. This is the reason why the phase space is made up of the diad of (position, momentum).

In its most general form we have the **Hamilton-Jacobi-Bellman equation**: [1]

$$\boxed{\text{Hamilton-Jacobi-Bellman}} \qquad 0 = \frac{\partial U^*}{\partial t} + \min_u H \tag{36}$$

All these "physical" ideas flow automatically from our definition of **rewards**, without the need to introduce them artificially. But these ideas seem not immediately useful to our project, unless we are to explore **continuous-time** models.

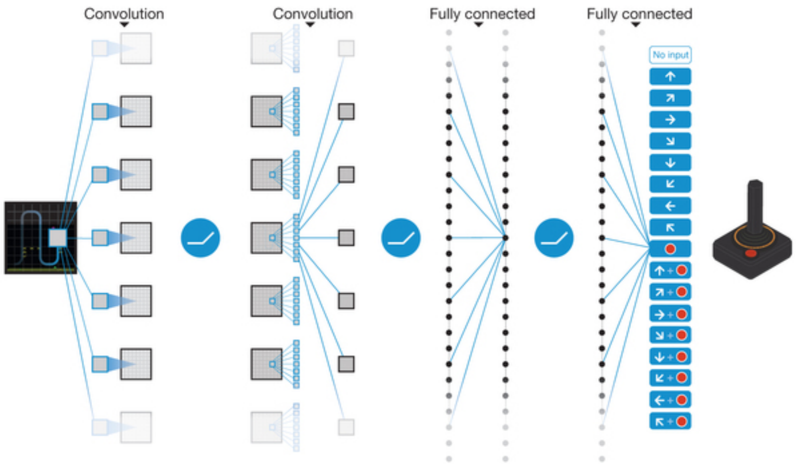
[1] To digress a bit, this equation is also analogous to the **Schrödinger equation** in quantum mechanics:

$$i\hbar \frac{\partial}{\partial t} \Psi(x, t) = \left[V(x, t) + \frac{-\hbar^2}{2\mu} \nabla^2 \right] \Psi(x, t). \tag{35}$$

where Ψ is analogous to our U (perhaps Ψ is something that nature wants to optimize?)

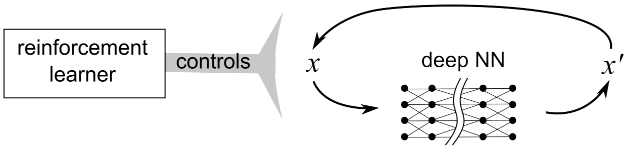
3.4 Deep learning

The combination of deep learning with reinforcement learning, ie deep reinforcement learning (DRL), is very powerful. For example, DRL is able to play Atari games to human levels [2]. Their architecture is depicted in this diagram:



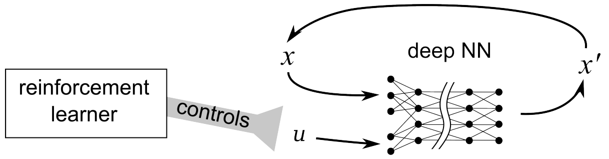
(37)

Recall our main architecture:



(38)

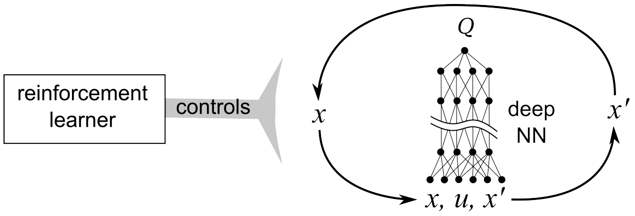
The "transition function" neural network can also take an **action** parameter u:



(39)

But such a neural network requires a novel learning algorithm that is **reward-driven** rather than the traditionally **error-driven** back-propagation. Such an algorithm is difficult to design because it cannot rely on old-fashioned gradient descent.

One solution that avoids the difficulty is to make the neural network compute the Q-value:



(40)

This means that we can compute Q for any transition $x \overset{u}{\mapsto} x'$. During the "action" (ie, "thinking") stage, we hold x fixed, and search for (u, x') that maximizes Q; This can be done by **stochastic gradient descent**. During the "learning" stage, we are given certain transitions (x, u, x') and we train the neural network to adjust Q via standard **Bellman update**.

Acknowledgement

In a forum discussion with Ben Goertzel dated 25 June 2014 on the AGI mailing-list: (agi@listbox.com), YKY asked: Why bother with neural networks, which typically require many neurons to encode data, when logic-based AI can represent a proposition with just a few symbols? Ben's insight is that neural networks are capable of learning their own representations, and their learning algorithms are relatively speaking much faster. We have been working on ”neo-classical” logic-based AI for a long time, and began to realize that inductive learning in logic (based on combinatorial search in a symbolic space) is perhaps *the bottleneck* in the entire logic-based paradigm. So we try to look for alternatives that might learn faster, though we would still emphasize that logic-based AI remains a viable approach to AGI.

References

1. Lo. Dynamical system identification by recurrent multilayer perceptrons. *Proceedings of the 1993 World Congress on Neural Networks*, 1993.
2. Mnih, Kavukcuoglu, Silver, Graves, Antonoglou, Wierstra, and Riedmiller. Playing atari with deep reinforcement learning. *arXiv:1312.5602 [cs.LG]*, 2013.
3. Siegelmann and Sontag. Turing computability with neural nets. *Applied Mathematics Letters, vol 4, p77-80*, 1991.
4. Weston, Chopra, and Bordes. Memory networks. *ICLR (also arXiv)*, 2015.